AD_____

Award Number: W81XWH-06-1-0057

TITLE: Identification and Characterization of an X-Linked Familial Prostate Cancer Gene

PRINCIPAL INVESTIGATOR: Brian Yaspan, Ph.D.

CONTRACTING ORGANIZATION: Vanderbilt University Medical Center
Nashville TN 37203-6917

REPORT DATE: November 2007

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01/11/07	Annual Summary	24 Oct 06 – 23 Oct 07

4. TITLE AND SUBTITLE
Identification and Characterization of an X-Linked Familial Prostate Cancer Gene

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0057

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Brian Yaspan, Ph.D.

E-Mail: brian.l.yaspan@vanderbilt.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Vanderbilt University Medical Center
Nashville TN 37203-6917

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Prostate cancer is the most commonly diagnosed non-skin malignancy in males, with as many as one in 5 males living in developed nations being diagnosed with prostate cancer in their lifetime. Despite the medical significance of prostate cancer our understanding of predisposition and progression in the disease remains rudimentary. Prostate cancer is estimated to have the largest heritable component of all common cancers. We will explicitly characterize ancestral versions of a gene region originally implicated in prostate cancer through study of families with multiple cases of prostate cancer (HPC) to enable us to test the hypothesis that a common disease-predisposing genetic mutation conferring modest risk is shared among present-day prostate cancer cases in the broader population by virtue of inheritance from an ancient founder. We hope the findings of this proposal will offer a promising inroad for predicting disease predisposition, for tailoring the most effective current therapy to each individual patient, and for developing rational new therapies.

15. SUBJECT TERMS
None provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	9	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

There is a significant heritable component of prostate cancer. Increased familial relative risk is observed across multiple populations. Male first degree relatives of prostate cancer patients have a two- to three-fold increased risk. Segregation analyses support genetic rather than shared environmental risk. Twin cancer concordance studies reveal a higher heritable risk for prostate cancer than for any other common cancer. Additional epidemiological studies have been consistent with X-linked transmission, identifying higher risk for a man with an affected brother relative to one with an affected father. Despite the overwhelming genetic predisposition evidence, the identification of prostate cancer susceptibility genes has been difficult. Linkage studies have resulted in the identification of several loci difficult to confirm across study populations. However, summary studies of genome-wide scans for prostate cancer susceptibility loci in general confirm two loci, HPC-1 and HPC-X.

Our study seeks to identify a candidate gene or genes conferring prostate cancer susceptibility at locus HPC-X in a US Caucasian study population. We hypothesize that a gene or genes at HPC-X harbor common moderate-penetrance variants predisposing to prostate cancer. We looked at shared haplotypes in founder populations and found two intervals likely to harbor prostate cancer susceptibility genes. We have chosen to first focus on one interval at locus HPC-X (termed HPC-X region A) due to shared haplotype association evidence in the founder populations of Finland, Iceland and Ashkenazim.

Body

Accomplishments

Tasks 1-3 have been altered to reflect the current state of the project.

Task 1. To identify and genotype all common polymorphism in our study population at potential genes of the candidate interval (HPC-X Region A). (Months 1-12):

a. Perform *de novo* SNP discovery at predicted or known genes and derive a set of survey SNPs spanning the HPC-X locus and a density of 3-5 kb from dbSNP
b. Genotype a subset of the study population for all SNPs in 1a
c. Analyze genotypes to determine genetic architecture of HPC-X

Task 1a-1c has been completed.

To derive a set of survey SNPs spanning the interval at HPC-X we assayed SNPs found in dbSNP for polymorphism in a subset of 40 prostate cancer cases from the training dataset. Out of 415 SNPs culled from database entries, we identified 194 as polymorphic and assayable in our study population. To augment this set of SNPs, we undertook *de novo* SNP discovery in the same subset of 40 prostate cancer cases at known and predicted genes in the region, identified in **Figure 1**. In addition to known

genes *SPANXC* and *LDOC1*, custom software identified a coding region containing homology to *RPL44* and a pseudogene containing homology to *RBMX2*. *De novo* SNP discovery at these four features resulted in 52 additional SNPs for a total of 246 SNPs. Linkage disequilibrium (LD) patterns at HPC-X Region A for these 246 SNPs typed in a subset of 141 controls of our training population are presented in **Figure 1**. Four major blocks of LD are apparent. Block "A" contains all four known and predicted genes.

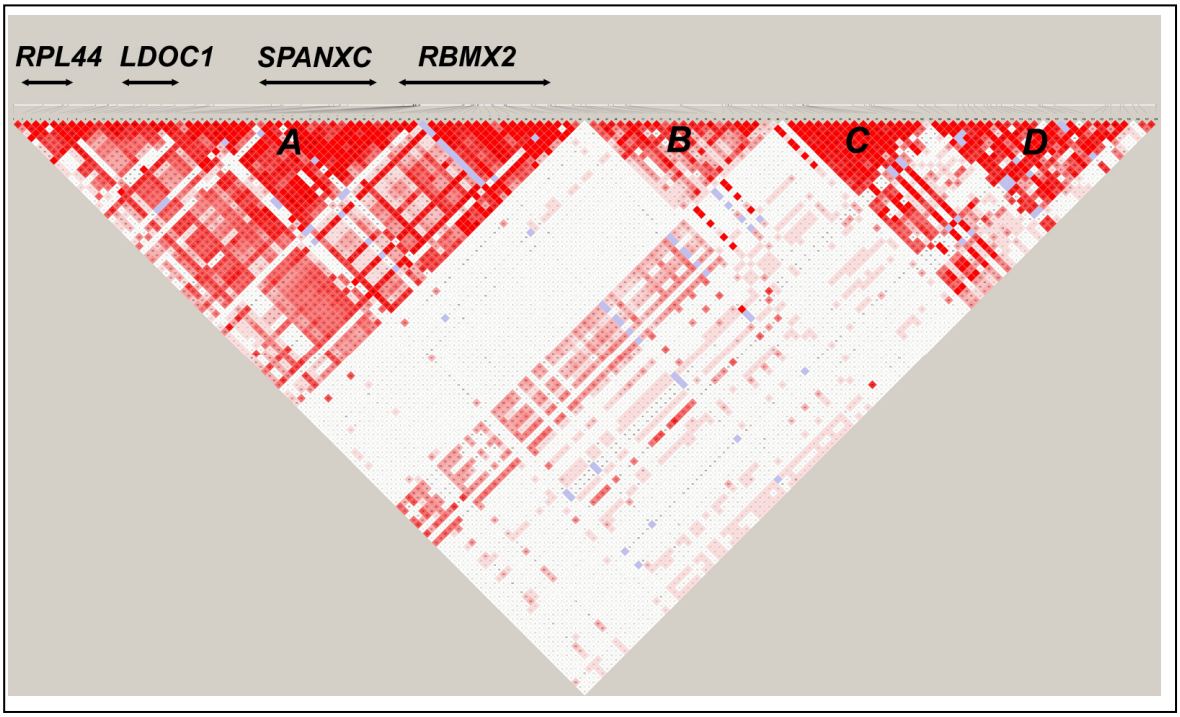

Figure 1: Linkage disequilibrium (LD) patterns at HPC-X Region A for 128 tagging SNPs for 141 controls. SNPs encompassing genes and 10 kb to each flank are positioned at top of the figure. Four major blocks are observed and marked as A, B, C and D. Red, D' = 1 (LOD ≥ 2); blue, D'= 1 (LOD < 2); pink, D' < 1 (LOD ≥2); white, D' < 1 (LOD < 2).

Task 2. To determine a set of tagging SNPs across our candidate interval at HPC-X, to genotype them in the training dataset and to test for significant association with risk of prostate cancer

a. Determine a set of tagging SNPs across our candidate interval at HPC-X from the 246 SNPs typed in *Task 1*
b. Genotype tagging SNPs in the remainder of the training dataset population
c. Using single allele and sliding window haplotype analysis, determine haplotype windows statistically associated with risk of prostate cancer

Task 2 a-c has been completed this year.

Using LDSelect, we have determined a set of 128 tagging SNPs ($r^2 > 0.9$) and typed them in the remainder of our training dataset (N = 292 cases, 292 age-matched

controls). All windows of statistical significance conferring risk of prostate cancer (nominal $P \leq 0.05$) within the candidate interval were from four distinct regions and are seen in **Table 1.** Regions are listed from 5' to 3' across the candidate interval. Individual SNPs in each haplotype are identified with an internal seven-digit code. Case and control frequencies are seen for the haplotype window with the most significant *P* value, colored in black. All other haplotype windows showing statistical significance are colored in grey.

Task 3. To confirm associated prostate cancer gene variants in a second study population, and to extend investigation in an African American Study population. (Months 24-36)

a. Ascertainment of independent population to be used as a test dataset
b. Confirm or refute areas of statistical association from Task 2 in test dataset
c. Extend findings into an African American study population, currently under ascertainment

Task 3 a, b has been completed this year. Task 3c is in progress.

We have recently ascertained an independent test dataset of 215 prostate cancer probands with a family history of disease and 215 age-matched controls. We used this dataset to confirm or refute statistical associations seen in the training dataset. We identified haplotype tagging SNPs (htSNP) for each of the four candidate risk haplotypes encompassing the entirety of each associated region and tested for association with risk of prostate cancer in our test dataset (**Table 2**). Case and control frequencies differ from those reported in Table 1 due to the use of only htSNPs to define the haplotype, allowing inclusion of some samples previously dropped from analysis due to missing data. Haplotype 3 was statistically significant within the test population (6.6% of cases, 2.5% of controls; $P = 0.04$). This haplotype spans areas "A" and "B" as seen in Figure 1, identifying a recombination hotspot. However, this association does not stand after Bonferroni correction for multiple testing bias given the number of individual tests performed in the test dataset (approximate adjusted $P = 0.12$).

We currently do not have sufficient power in our African American dataset and as a result, we are actively increasing our ascertainment efforts. To date we have 150 African American cases and 87 African American controls.

Table 1. Significant Risk Haplotypes at HPC-X Region A - Training Dataset					
Haplotype	**Significant Haplotype Windows**	**Allele**	**Cases,**	**Controls,**	***P***
Haplotype 1	2909545	A	92 (39.8)	63 (27.3)	0.003
	2918393	C			
	2924155	G			
	2924499	T			
	2925845	G			
	2934172	A			
	2937803	A			
Haplotype 2	1263092	G	15 (8.6)	5 (2.9)	0.021
	3046732	A			
	1206331	G			
	1260272	A			
	3047533	T			
	1259699	C			
	1259001	A			
	1258867	A			
	3048960	A			
	3050562	G			
	3052976	A			
Haplotype 3	3056703	T	15 (6.9)	3 (1.4)	0.003
	3057136	T			
	3069739	C			
	3064694	A			
	3065097	C			
	3078481	A			
	3069855	T			
	3070491	C			
	3083169	C			
	3084363	A			
	3077151	C			
	3077417	C			
	3091977	A			
Haplotype 4	3145284	T	21 (8.9)	9 (3.8)	0.024
	3146375	T			
	3138940	G			
	3140443	A			
	3141049	G			
	3143286	A			
	3156248	C			
	3146147	T			
	3158651	A			
	3146512	G			
	3147351	C			
	3160094	A			
	3173563	C			

Table 2. χ^2 Test for Association Training and Test Dataset Significant Haplotypes

Haplotype	Marker	Allele	Training			Test			Combined		
			Cases,n(%)	Controls,n(%)	*P*	Cases,n(%)	Controls,n(%)	*P*	Cases,n(%)	Controls,n(%)	*P*
Haplotype 1	2918393	C	95 (38.6)	71 (28.9)	0.023	73 (39.3)	79 (42.5)	0.536			
	2924155	G									
	2924499	T									
	2925845	G									
	2934172	A									
Haplotype 2	3046732	A	19 (6.6)	9 (3.1)	0.062	10 (4.9)	9 (4.4)	0.767			
Haplotype 3	3069739	C	18 (6.8)	7 (2.6)	0.02	13 (6.6)	5 (2.5)	0.04	31 (6.7)	12 (2.6)	0.003
	3078481	A									
	3069855	T									
	3083169	C									
	3084363	A									
	3091977	A									
Haplotype 4	3140443	A	22 (7.7)	14 (4.9)	0.168						
	3156248	C									

Key Research Accomplishments

1. Ascertainment of a US Caucasian study population with statistical power to detect common variants that may predispose prostate cancer risk.
2. *De novo* SNP discovery leading to discovery of 52 unpublished SNPs (at time of discovery) in a US Caucasian population. At time of writing 32 are still unpublished and have been submitted to dbSNP.
3. Identification of a 22.8 kb area spanning a recombination hotspot tagged by 6 htSNPs associated with risk of prostate cancer in both test and training datasets.

Reportable Outcomes

Investigation of a candidate locus at HPC-X in familial prostate cancer
Poster presentation at IMPaCT meeting, Hyatt Regency Atlanta 2007
Yaspan B, McReynolds K, Elmore JB, Breyer J, Bradley K, Smith JR

No association with risk of prostate cancer for *LDOC1* and *SPANX-C* candidate genes within the HPC-X locus in a US Caucasian study population
Poster presentation at the American Society of Human Genetics Meeting, New Orleans, LA 2006
Yaspan B, Elmore JB, Breyer J, Bradley K, McReynolds K, Smith JR

Conclusions

Over the past two years, we have been systematically dissecting HPC-X to uncover the variant or variants responsible for its association with risk of prostate cancer. Starting with region A, whose boundaries we identified through shared haplotype analysis of founder populations, we have identified one haplotype tagged by 6 htSNP spanning a 22.8 kb region pinpointing a recombination hotspot associated with risk of prostate cancer. This haplotype was statistically significant for risk of prostate cancer in both our test and training datasets.

Future Directions

In the next year we will begin a HapMap based analysis over the entirety of HPC-X. Recently, several reports have identified and confirmed multiple variants predisposing to risk of prostate cancer at chromosome 8q24. We believe there could be multiple variants at HPC-X which are analogous to 8q24 and are focusing on the entirety of the locus. We intend to utilize the International HapMap for selecting htSNPs across HPC-X for typing in our population by looking at the htSNP ($r^2 > 0.80$) selected for the CEPH population (Utah residents with ancestry from northern and western Europe). We will select htSNP for genes identified in public databases and potential coding regions identified using custom bioinformatics software. We will then genotype this set of htSNP in our study population.